Analyzing Book-Related Features to Recommend Books for Emergent Readers

Maria Soledad Pera
Department of Computer Science
Boise State University
Boise, Idaho, U.S.A.
solepera@boisestate.edu

Yiu-Kai Ng
Computer Science Department
Brigham Young University
Provo, Utah, U.S.A.
ng@compsci.byu.edu

ABSTRACT

We recognize that emergent literacy forms a foundation upon which children will gage their future reading.[1] It is imperative to motivate young readers to read by offering them appealing books to read so that they can enjoy reading and gradually establish a reading habit during their formative years that can aid in promoting their good reading habits. However, with the huge volume of existing and newly-published books, it is a challenge for parents/educators (young readers, respectively) to find the right ones that match children's interests and their readability levels. In response to the needs, we have developed K3Rec, a recommender which applies a multi-dimensional approach to suggest books that simultaneously match the *interests/preferences* and *reading abilities* of emergent (i.e., K-3) readers. K3Rec considers the grade levels, contents, illustrations, and topics, besides using special properties, such as length and writing style, to distinguish K-3 books from other books targeting more mature readers. K3Rec is novel, since it adopts an unsupervised strategy to suggest books for K-3 readers which does not rely on the existence of personal social media data, such as personal tags and ratings, that are seldom, if ever, created by emergent readers. Furthermore, unlike existing book recommenders, K3Rec explicitly analyzes book illustrations, which is of special significance for emergent readers, since illustrations assist these readers in understanding the contents of books. K3Rec focuses on a niche group of readers that has not been explicitly targeted by existing book recommenders. Empirical studies conducted using data from BiblioNasium.com and Amazon's Mechanical Turk have verified the effectiveness of K3Rec in making book recommendations for emergent readers.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Information filtering

[1]http://www.deafed.net/publisheddocs/sub/9807kle.htm

HT'15, September 01 - 04, 2015, Guzelyurt, TRNC, Cyprus
© 2015 ACM. ISBN 978-1-4503-3395-5/15/09 ...$15.00.
DOI: http://dx.doi.org/10.1145/2700171.2791037.

Keywords

Book recommendations; emergent readers; K-3

1. INTRODUCTION

Reading is an activity performed on a daily basis: from reading news articles and books to cereal boxes and street signs. According to the National Institute of Child Health and Human Development, "reading is the single most important skill necessary for a happy, productive, and successful life",[2] which is the reason why focusing on *emergent* (or early) *reading* that refers to the knowledge, skills, and dispositions acquired in reading (and writing) in primary school grades prior to and up till the 3^{rd} grade [23], is particularly significant. As stated in [27], learning to read is a key milestone for children living in a literate society, specially given that reading provides the foundation for children's academic success. A recent study [4] highlights the fact that children who "do not read proficiently by the end of third grade are four times more likely to leave school without a diploma than proficient readers." The results of the study correlate with earlier statistics [11] which confirm that 88% of children who are poor readers by the end of the first grade remain so by the end of the fourth grade. Moreover, young readers who successfully learn to read in the early primary years of school will more likely be prepared to read for pleasure and learning in the future [18]. The aforementioned findings constitute the essence of encouraging good reading habits early on. Identifying books appealing to emergent readers (i.e., readers up till the 3^{rd} grade), however, can be challenging, given the amount of books made available on a regular basis that address a diversity of topics and target readers at different reading levels. It is essential to provide emergent readers with reading materials matching their preferences/interests and reading abilities, since exposing young readers to materials that are either too easy/difficult to understand or involving unappealing topics could diminish their interest in reading [1].

In the quest for locating print materials (especially books) which can help develop/improve the reading skills of K-3 readers, parents, educators, and young readers can turn to online book recommendation systems which suggest books of potential interest. Unfortunately, existing book recommenders [10, 25] require user-defined information, such as tags, ratings, connections, and accessing patterns, to make suggestions for the respective individuals. Personal information of K-3 users, however, may not exist owing to the lack

[2]http://www.ksl.com/?sid=15431484

of online social networking sites targeting K-3 users or may not be publicly accessible due to the ethical obligation of everyone to respect the online privacy of children. Moreover, majority of these recommenders fail to explicitly consider (i) the reading ability of a reader, which is necessary in making recommendations for readers with diverse reading skills [26], and/or (ii) unique characteristics that distinguish books targeting emergent, as opposite to advanced, readers [22].

To solve the problems in suggesting books for emergent readers, we have developed *K3Rec*, an unsupervised books recommender, which facilitates one of the tasks undertaken by parents/educators/young readers on a daily basis: to identify books that help improve their reading abilities of K-3 readers. K3Rec applies a multi-dimensional analysis on a book known to be of interest to a reader R and identifies other relevant books from existing book repositories, such as OpenLibrary.org, that match (to a degree) the preferences and reading ability of R. While the criteria that dictate an appropriate K-3 book are determined using a number of pre-defined features that commonly apply to "good" books targeting emergent readers [22], its correlation with the preferences and reading ability of R is analyzed by conducting an in-depth examination on a brief description of its content, pictorial perspectives, reading level, and topics as defined based on Library of Congress Subject Headings.

K3Rec is a novel recommender that exclusively targets emergent readers, an audience who has not been catered by existing recommendation systems. K3Rec is a self-reliant recommender which, unlike others, does not rely on the availability of *personal information* about its users to make book suggestions. Instead, K3Rec takes advantage of book metadata, which are either readily and freely available from reputable online sources, such as the Library of Congress (catalog.loc.gov), or inferred from user-defined metadata, such as book reviews and book ratings, that are publicly accessible online from popular book-related websites, e.g., Google Books (http://books.google.com/) and Amazon.com. K3Rec is unique, since it explicitly considers one of the most distinguishable aspects of books for emergent readers [9, 22]—their illustrations—by employing OpenCV (opencv.org) an open source computer vision/machine learning software.

K3Rec is designed for solving the *information overload* problem while minimizing the *time* and *efforts* imposed on parents/educators/young readers in discovering unknown, but suitable, books for pleasure reading or knowledge acquisition. The current implementation of K3Rec is tailored towards recommending books written in English and classified based on the K-12[3] grade level system. K3Rec, however, can be easily adopted to make suggestions based on diverse grade-level scales and in languages other than English.

The remaining of this paper is organized as follows. In Section 2, we discuss existing recommenders that have been used for identifying books for individual readers, including young readers. In Sections 3, we introduce K3Rec and its overall design methodology. In Section 4, we present the results of the empirical studies on K3Rec conducted to assess its performance. In Section 5, we give a concluding remark and present directions for future work on K3Rec.

[3]K-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 12^{th} grades.

2. RELATED WORK

To the best of our knowledge, there is no existing book recommendation system developed specifically for emergent readers. At present, parents/educators/young readers often rely on existing book websites, including, but not limited to, ARbookfind.com, Kidsread.com, Scholastic.com, and WorldCat.org, which offer different tools to search for books in various domains. These sites, however, either (i) supply (read-alike) non-personalized booklists [8], (ii) require a particular topic/subject area of interest to be selected from a pre-defined list,[4] which limits the themes of books that can be obtained from the sites, (iii) offer reading choices grouped by age/grade ranges,[5] which is undesirable, since readers in the same grade or age group might not reach the same reading level, or (iv) allow users to create keyword queries to specify their information needs, which often yield an overwhelming volume of items to choose from and impose an additional burden on users to sort though. Unlike the aforementioned websites, K3Rec eliminates their constraints imposed in locating books, which enhances the process in finding books relevant to the information needs of emergent readers and at a reading level appropriate for the readers.

Even though there are no book recommenders for emergent readers, 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 [14], whereas Yang et al. [28] analyze users' access logs to infer the users' preferences and apply the traditional collaborative-filtering (CF) strategy to make book recommendations. The authors in [10] combine CF and social tags to capture the content of books for making recommendations. Sieg et al. [25], on the other hand, 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. BReK12 [19], which is based on content and readability analysis, relies heavily on the availability of bookmarking information offered by social bookmarking sites to suggest K-12 books. Unlike K3Rec, these recommenders require (i) historical data on the users in the form of ratings and bookmarking information, which may not always be accessible, or (ii) an ontology, which can be labor-intensive and time-consuming to construct. In addition, none 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 K3Rec is not a recommender for direct learning, its design goal is to enhance reading selections for emergent readers by locating suitable books among the overwhelming number of choices available these days. (For an in-depth description of existing recommenders in the educational domain, see [16].)

3. OUR PROPOSED RECOMMENDER

In making book suggestions for a K-3 reader R, K3Rec first analyzes a given book B known to have been read by R and identifies books that are compatible with the readability level of R (detailed in Section 3.1). These books are treated as *candidate books* to be considered for recommendation. Candidate books are selected among the books

[4]http://www.readingrockets.org/books/booksbytheme
[5]http://goo.gl/78X7i6

available at one of the (online) book repositories, which include, but are not limited to, (i) *reputable websites*, such as OpenLibrary.org or WorldCat.org, which are two of the largest online library catalogs, (ii) *school/public libraries*, and (iii) book-related *bookmarking sites*, such as Biblionasium.com, which is a website that encourages reading among children/teenagers. K3Rec computes a ranking score (in Section 3.3) for each candidate book CB, which captures not only the degree of *context closeness* of CB and B, but also the *desired properties* of books for emergent readers that apply to CB for R based on the analysis of multiple book-related features (presented in Section 3.2).

3.1 Identifying Candidate Recommendations

One of the design goals of K3Rec is to suggest books that its readers can comprehend. It is imperative for K3Rec to locate books with grade levels suitable for a reader R, since "reading for understanding cannot take place unless the words in the text are accurately and efficiently decoded" [17]. K3Rec determines the readability level of R based on the grade level of a given book B, which is computed using TRoLL [20], a regression-based readability prediction tool. Unlike existing popular readability-level prediction formulas/tools, such as Flesch-Kincaid, Lexile Framework, and ATOS (discussed in details in [2]), TRoLL computes the grade level of a book using metadata on books publicly accessible from reputable online sources, even in the absence of book excerpts. Hence, TRoLL is not constrained by the availability of sample text of a book, which is not always freely accessible due to copyright laws. Experimental results [20] show that TRoLL is highly accurate in predicting the grade levels of K-12 books and outperforms other existing readability formulas/tools, such as Flesch-Kincaid and Accelerated Reader (AR), which rely on books excerpts.

Based on the readability level of a reader R through B, K3Rec applies Equation 1 to determine the set of candidate books considered for recommendation to R.

$$SCB(B) = \{CB \mid CB \in Rep \wedge RL(CB) \in [RL(B) \pm 0.25]\} \quad (1)$$

where CB is a candidate book available at a book repository Rep and $RL(CB)$ ($RL(B)$, respectively) is the grade level of CB (B, respectively) determined by TRoLL. By selecting books within *half a grade*[6] of the grade level of B, K3Rec considers books for recommendation within an appropriate level of (text) complexity for R based on the grade level of B that R is interested in the past.

Example 1. Consider a reader R_A, who has read the books "If you give a pig a party" by Laura Numeroff and "Fancy Nancy" Nancy O'Connor. Using TRoLL, K3Rec determines that the readability levels of these books are 1.10 and 1.40, respectively. Based on this information, K3Rec establishes 1.25 ($= \frac{1.10+1.40}{2}$) as the readability level of R_A. Using Equation 1, K3Rec generates a set of candidate books which includes books from the BiblioNasium dataset (introduced in Section 4.1) with readability levels between 1.0 and 1.5. Consequently, books such as "The paperboy" by Dav Pilkey, "If you give a mouse a cookie" by Laura Numeroff, and "Cat and dog" by Else Holmelund with readability levels 1.15, 1.3, and 1.45, determined by TRoLL, respectively, are considered as candidate books to be considered for recommendations for R_A, since they can be read and comprehended by R_A. Furthermore, books such as "Harry, the Poisonous Centipede" by Lynne Reid Banks and "Football Genius" by Tim Green with readability levels 0.25 and 2.2, computed by TRoLL, respectively, are excluded from the candidate set, since they are too easy and too challenging for R_A, respectively. □

3.2 Book-Related Feature Analysis

K3Rec suggests relevant books not only readers are interested in, but also they can comprehend. This is accomplished by examining candidate books (determined using Equation 1) using diverse publicly accessible book metadata to analyze (i) *book contents* appealing to R (in Section 3.2.1), (ii) the *type of illustrations* of interest to R (in Section 3.2.2), and (iii) the general *traits* applied to CB that are significant factors to be considered for books targeting emergent readers (in Sections 3.2.3 - 3.2.6).

3.2.1 Content Analysis

K3Rec analyzes the content description of CB, which can be extracted from reputable book-related websites, such as Amazon.com and the Library of Congress, to determine the degree to which CB addresses subject matters that are appealing to R based on the overview of B. As shown in Equation 2, K3Rec computes the *content similarity* score between CB and B, denoted $CSim(CB, B)$, based on the "bag-of-words"[7] representation of the description of CB and B. $CSim(CB, B)$ considers the word-correlation factor (wcf) [13] of each word in the description of B with respect to each word in the description of CB, and prioritizes candidate books based on their degree of shared content with B. 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(.com) documents. K3Rec relies on word-correlation factors, instead of similarity measures [3] based on WordNet(.pricenton.edu), since it has been empirically verified that the former correlates with human assessments on word similarity more accurately than the latter [19].

$$CSim(CB, B) = \frac{\sum_{i=1}^{n} Min\{\sum_{j=1}^{m} wcf(B_i, CB_j), 1\}}{n} \quad (2)$$

where n (m, respectively) is the number of distinct words in the description of B (CB, respectively), B_i (CB_j, respectively) is a word in the description of B (CB, respectively), and $wcf(B_i, CB_j)$ is the correlation factor, i.e., degree of similarity, of B_i and CB_j in the word-correlation matrix.

The Min function in Equation 2 imposes a constraint on summing up the correlation factors of words in the description of CB and B. Even if a word in the description of B (i) matches exactly one of the words in CB and (ii) is similar to some of the remaining words in CB, which yields a value greater than 1.0, K3Rec limits the sum of their similarity measure to 1.0, which is the word-correlation factor of an *exact* match. This constraint ensures that if B contains a

[6] We have empirically verified that by selecting 0.25 as a threshold in Equation 1, the overall processing time of K3Rec is shortened, without significantly affecting its accuracy.

[7] From now on, unless stated otherwise, "word" refers to non-stop, stemmed word.

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 B with respect to CB. Words in the brief overview of CB that are similar to most of the words in B should yield a greater $CSim$ value than the $CSim$ value of words in the description of CB that are similar to only one dominant word in B.

3.2.2 Illustration-Based Analysis

One of the features commonly associated with a book for emergent readers is its illustrations. Since illustrations play an important role in "directly encouraging children's emergent literacy development" [12], it is imperative for K3Rec to consider book illustrations as part of its recommendation process. Similar to the textual content of a book, its illustrations are not always freely accessible due to copyright laws. However, there are a number of websites that offer API access to book covers, such as LibraryThing.com and Google Books. K3Rec takes advantage of such resources and calculates $Isim(CB, B)$, a score that reflects the degree of resemblance between the illustrations as shown on the book covers of CB and B. K3Rec prioritizes candidate books partially based on the illustrations as shown in their covers with similar images to the book known to be appealing to R.

It is not an easy task, however, to compute $Isim(CB, B)$, given that the similarity between images is based on accurately identifying the same (or similar) object(s) or scene(s) even if they are presented under different imaging conditions, such as viewpoint changes, image blur, and illumination changes [15]. To facilitate the task of determining the degree of similarity between any two book covers, K3Rec applies the Open Source Computer Vision (OpenCV) library. Given any two images, i.e., the book covers of CB and B, OpenCV models them as matrices of multiple image features. These matrices are then compared to determine $Isim(CB, B)$ that quantify the degree of resemblance between the two images.

Example 2. Consider the book covers as shown in Figure 1, which correspond to "Don't let the pigeon drive the bus" by Mo Willems ($Book_A$), " The pigeon finds a hot dog" by Mo Willems ($Book_B$), and "Pat the bunny" by Dorothy Kunhardt ($Book_C$). Using OpenCV, K3Rec determines that $Isim(Book_A, Book_B)$ is higher than $Isim(Book_A, Book_C)$. This is anticipated, since although the covers of $Book_A$ and $Book_C$ share very similar background colors, the covers of $Book_A$ and $Book_B$ share similar images, i.e., the pigeons and dialogue bubbles. Based on the computed $Isim$ scores, K3Rec prioritizes $Book_B$ over $Book_C$ in making suggestions for a reader given his/her interest in $Book_A$. □

3.2.3 Topical Analysis

Besides considering the relatedness of CB and B based on their content representations and illustrations, K3Rec examines *topical information* of CB to determine its suitability for R. This analysis is based on Library of Congress Subject Headings (LCSH) assigned to CB by professional cataloguers. LCSH, which is a de facto universal controlled vocabulary, constitutes the largest general indexing vocabulary in the English language [29]. LCSH, which are *terms* or *phrases* that denote concepts, events, or names, are used by librarians to categorize and index books according to their themes. Examples of LCSH include "Fairy tales" and "Fear of the dark–Fiction".

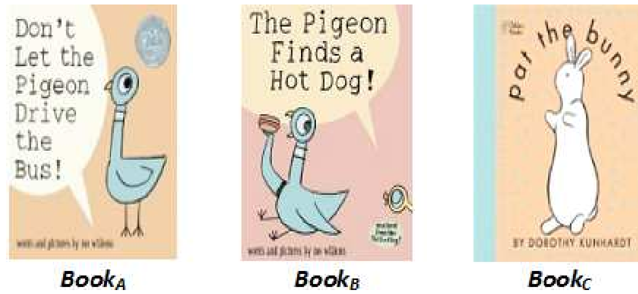

Figure 1: Sample book covers

Features derived from the LCSH of CB, which are publicly accessible from the Library of Congress, include their (i) total count and (ii) associated grade levels.

Total Count of LCSH. K3Rec considers the *count* of LCSH assigned to CB, since books that are *more difficult* to comprehend are often assigned *more* LCSH[8]. The degree of difficulty in comprehending CB (based on its subjects), denoted $Diff(CB)$, is computed by K3Rec using Equation 3, which penalizes candidate books that have been assigned more LCSH than other books in the set of candidate books considered for recommendation, since the *lower* the number of LCSH assigned to CB, the *more likely* the audience targeted by CB are emergent readers.

$$Diff(CB) = \frac{1}{|LCSH_{CB}|} \tag{3}$$

where $LCSH_{CB}$ is the set of LCSH assigned to CB and $|LCSH_{CB}|$ denotes the size of $LCSH_{CB}$.

LCSH and Grade Levels. Besides using the *count* of LCSH, K3Rec also considers the grade levels associated with LCSH assigned to a (candidate) book. Using Equation 4, K3Rec determines the proportion of LCSH of CB that are associated with grade levels similar to the grade (i.e., readability) level of R (through book B). K3Rec favors candidate books that address subjects suitable to the reading level of R, which is one of the major goals of K3Rec, i.e., suggesting books tailored to the reading abilities of individual readers.

$$LC(CB, B) = \frac{\sum_{j=1}^{|LCSH_{CB}|} isSuitable(CB_j, RL(B))}{|LCSH_{CB}|} \tag{4}$$

where $LCSH_{CB}$ and $|LCSH_{CB}|$ are defined in Equation 3, CB_j is the j^{th} LCSH in $LCSH_{CB}$, and $isSuitable(CB_j, RL(B))$ is a function that returns "1" if the grade level of CB_j is within a quarter of $RL(B)$ (as defined in Equation 1), and is "0" otherwise. Note that the grade level associated with a given LCSH is determined based on the mapping between grade levels and LCSH defined in [20]. (See Table 1 for sample mappings between LCSH and their grade levels, where "1.5" indicates that the corresponding LCSH, i.e., "Babar fictitious character," is often assigned to books between the first and second grade.)

[8]The authors in [20] have empirically verified the correlation between the number of LCSH assigned to K-12 books and their corresponding grade levels. The analysis in [20] has shown that the lower the number of LCSH assigned to a books is, the lower is the grade level defined for the book.

Table 1: Sample mapping between LCSH and grade levels

LCSH	Grade Level
Babar fictitious character	1.5
Bedtime fiction	1.8
Bedtime prayer	0.2
Dora the explorer fictitious character	0.8
Scary stories	2.8
Zoo-children-fiction	0.4

3.2.4 Book-Length Analysis

Another desired property of books for emergent readers is the *length*, i.e., the number of pages, of the books. As stated in [21], books for emergent readers are on an average of 32 pages in length. Relatively *short* books are preferred, since they can be read in one (or few) sittings, which offers their readers a sense of accomplishment in finishing a book.[9] K3Rec applies Equation 5 to measure the degree to which the length of CB is within the expected length of a book targeting emergent readers.

$$Len(CB) = \begin{cases} 1 & \text{if } Pages(CB) \leq 32 \\ \frac{1}{Pages(CB) - 32} & \text{otherwise} \end{cases} \quad (5)$$

where $Pages(CB)$ is the number of pages of CB, which can be obtained by accessing the publicly available catalog record for CB from the Library of Congress.

As shown in Equation 5, K3Rec imposes a penalization on books longer than 32 pages. This penalization is scaled to the number of pages of CB such that the *more* pages that exceed the average number of pages expected for a K-3 book, the *lower* the chance CB targets K-3 readers.

3.2.5 Writing Style-Based Analysis

Another characteristic often applied to books for emergent readers is the *simplicity* and *directness* of their texts [21]. Identifying the writing style of books, however, is non-trivial, given the lack of access (due to copyright laws) to sample text on books required to perform semantic/syntactic analysis. An alternative to gather this information is to turn to book metadata available at online sources, such as NoveList, which provide a description of the literary elements of a book. Literary elements are "elements of a book—whether definable or just understood—that make readers enjoy the book" [24]. These elements, which include characterization, frame, pacing, storyline, language and writing style, and tone, capture general traits of a book [5]. Access to these resources, however, requires a paid subscription. K3Rec relies on ABET [20] instead to obtain a description of the writing style of each candidate book CB.

ABET is a newly-developed, unsupervised tool that automatically generates a description of the literary elements 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 and Powell.com. By analyzing reviews, ABET determines diverse readers' opinions on a book based on terms (also known as appeal terms) that describe the corresponding literary elements (i.e., appeal factors) of the book. A sample of the appeal terms and appeal factors considered by ABET are included in Table 2.

[9]http://www.rif.org/documents/us/choosing_books.pdf

Table 2: Sample appeal terms associated with each of the appeal factors considered by ABET

Appeal Factors	Appeal Terms
Characterization	Believable, distant, dramatic
Frame	Bittersweet, contemporary, descriptive
Language and Writing Style	Candid, complex, conversational, extravagant, poetic, prosaic
Pacing	Easy, fast, slow
Special Topics	Addiction, bullying, violence
Storyline	Action-oriented, character-centered
Tone	Dark, happy, surreal

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 literary elements and terms used to describe them, which are based on typed dependency relations. It is natural for ABET to turn to typed dependencies, since they capture the *semantic connection*, i.e., association, between words in sentences. For this reason, the rules defined for ABET simply look for words in sentences that (directly or indirectly) describe the literary elements of a book, which are often the subjects or objects of sentences[10].

The rules introduced in [20] to extract a writing-style description for a book based on its corresponding reviews are defined in Table 3. These rules, which are used to generate descriptions of appeal factors, including writing style, are based on common writing patterns identified in book reviews and capture the semantic link between appeal factors and their corresponding terms that describe them. Consider the sentence S_A, "The words in the book are simple", and sentence S_B, "The author creates unmistakable, classic characters". In S_A the *subject* of the sentence, i.e., "words," is characterized as being "simple", whereas in S_B its *object*, i.e., "characters", is described as "classic". In these examples, it is clear that if the subject/object of a sentence is an appeal factor, then a word in the sentence that semantically describes, i.e., is directly linked to, the mentioned object/subject is often its descriptive keyword, i.e., appeal term. ABET captures these connection patterns using Rules 1 and 2 as defined in Table 3.

An appeal term can also be indirectly connected with an appeal factor in a sentence. Consider sentence S_C, "The characters portrayed are funny." "Funny" is *indirectly* related to the subject of S_C, i.e., "characters", through the word "portrayed". Using Rule 3, ABET examines pairs of grammatical relations that involve indirect connections among words. Next, consider S_D, "The writing is not direct". Based on Rule 1, ABET would mistakenly describe the appeal factor "Writing Style" using the keyword "direct." This exam-

[10]Despite being comprehensive, the taxonomy defined for ABET that enumerates appeal factors and appeal terms cannot account for every variation of appeal factors/terms that can be specified in readers' reviews. For example, a reviewer may refer to the "Storyline" of a book as "story" or "narrative", and (s)he may also use either "easy" or "simple" as the keyword that describes the "Writing Style" of a book. To handle these variations during the extraction process, ABET uses (stemmed) synonyms of each appeal factor/term, which can be identified using WordNet.

Table 3: Rules considered by ABET to extract writing-style descriptions in book reviews

Notations
$rel(A, B)$ is a *grammatical relation* between a *dominant*, i.e., *governor* or *head*, word (A) and a *subordinate*, i.e., *dependent* or *modifier*, word (B)
L_F, L_T, EL_F, and EL_T are the list of appeal factors, list of appeal terms, extended list of appeal factors, and extended list of appeal terms, respectively.
w_f is an appeal factor in L_F, and w_t is an appeal term in L_T
$w \rightsquigarrow w_f$ ($w \rightsquigarrow w_t$, respectively) denotes that w is a synonym of w_f (w_t, respectively)
$POS(w)$ is the part-of-speech tag of w which is a verb (adverb, respectively) if $POS(w)$ = "VB" ("RB", respectively)
Abbreviation: adv(erbial)mod(ifier), a(djectival)mod(ifier), c(lausal)comp(lement), d(irect)obj(ect), neg(ation modifier), nn (noun compound modifier), n(nominal)subj(ect), nsubjpass (passive nominal subject), prep(_*) (Prepositional modifier)
ABET only extracts a pair $< w_f, w_t >$ if w_t is in the corresponding vocabulary defined for w_f

Rule	Objective	Conditions	Identified Factors/Terms
1	To capture the written patterns based on a keyword, i.e., appeal term, that immediately precedes/ follows the subject or object of a sentence S, i.e., appeal factor	$A \in EL_T, B \in EL_F, rel \in$ {nn, nsubj} $A \rightsquigarrow w_t$	$B \rightsquigarrow w_f$
2		$A \in EL_F, B \in EL_T, rel \in$ {advmod, amod, prep_in, prep_about}	$A \rightsquigarrow w_f$ $B \rightsquigarrow w_t$
3	To identify an appeal term that qualifies its indirectly related appeal factor in S	$rel \in$ {nn, nsubj}, $B \in EL_F$, and $\exists rel_2(C, D) \in$ {amod, dep, ccomp}, $A = C, D \in EL_T$	$B \rightsquigarrow w_f$ $D \rightsquigarrow w_t$
4	To explicitly consider *negated* appeal terms in S	$B \in EL_F, rel \in$ {nn, nsubj}, $\exists neg(C, D)$, A $(= C)$ is an antonym of $\bar{A} \in EL_T, D$ is a negation term	$B \rightsquigarrow w_f$ $\bar{A} \rightsquigarrow w_t$

ple shows the necessity of examining pairs of grammatical relations in the presence of negated terms. ABET applies Rule 4, which identifies a negated term as a modifier of a keyword k and then extracts as the keyword description for the corresponding feature the antonym of k (if it is included in the vocabulary defined in ABET's taxonomy for the feature). Together, Rules 1 to 4 account for the most common written patterns for appeal factors/terms observed in reviews. These rules look for words in sentences that (directly or indirectly) describe the qualitative features of a book, which are often the subjects or objects of sentences. Rules 3 and 4 take precedence over Rules 1 and 2, since once a dependency in a sentence is used by either of the former rules, it cannot be considered by the latter ones.

It is important to note that the description of the writing style of CB determined by ABET involves not only the terms extracted from reviews on CB that describe the language and writing style of CB, but also their *frequency of occurrence*. The latter captures the relative *degree of significance* of a term in describing the writing style of CB based on reviewers' varied opinions expressed in their reviews.

Using the ABET-generated writing style description of CB, K3Rec applies Equation 6 to compute $WTS(CB)$, which quantifies the degree of *directness* and *simplicity* of (the textual content of) CB. The *higher* $WTS(CB)$ is, the *larger* the number of reviewers who describe the writing style of CB as simple/direct, which reflects the *more likely* that CB includes text expressed in a simple/direct manner, a criteria of books suitable for emergent readers.

$$WTS(CB) = \frac{\sum_{i=1}^{|WSDsc|} isDirect(WSDsc_i)}{\sum_{i=1}^{|WSDsc|} |WSDsc_i|} \quad (6)$$

where $WSDsc$ is the set of distinct terms in the ABET-generated writing style description of CB, $|WSDsc|$ is the size of $WSDsc$, $WSDsc_i$ is the i^{th} term in $WSDsc$,

- **Jayne Eyre** by Charlotte Brontë

Complex (8), passionate (6), simple (1), unusual (9), classic (6)

- **The Pigeon Finds a Hot Dog!** by Mo Willems

Simple (9), dramatic (1), direct (5), classic (1)

Figure 2: Example of ABET-generated writing style descriptions, where the number (in parentheses) indicates the *frequency* in which a term was used to describe the corresponding writing style of books in reviews

$|WSDsc_i|$ denotes the *frequency* in which $WSDsc_i$ appears in the ABET-generated writing style description of CB, and $isDirect(WSDsc_i)$ denotes the *frequency* of $WSDsc_i$ if the term is "simple" or "direct," and is "0" otherwise.

Example 3. Consider the ABET-generated descriptions of the writing style of the books "Jane Eyre" by Charlotte Bronte and "The pigeon finds a hot dog!" by Mo Willems as shown in Figure 2. WTS("Jane Eyre") = $\frac{1}{30}$ = 0.03, whereas WTS("The pigeon finds a hot dog!") = $\frac{14}{16}$ = 0.88. Based on the WTS scores, K3Rec favors the latter for recommendation, which is anticipated, since the latter is indeed a book for emergent readers. □

3.2.6 Rating Assessment

Another feature considered by K3Rec in estimating the degree of appealing of CB is its *rating*. As product ratings capture an independent measure of the quality of a product based on the opinions of a number of appraisers who are familiar with the product [7], it is natural for K3Rec to prioritize books that have been assigned a *high* rating. The rating score of CB, denoted $Rate(CB)$, is extracted from

Google Books' API[11] which is the *average* of the ratings given to CB by Google Book users.

Note that even though K3Rec turns to the "wisdom of crowds" for another appeal measure, i.e., rating, on candidate books, it is completely different from the strategies employed by existing book recommenders [28]. The latter rely on the availability of *personal ratings* assigned to books by an individual user (to reflect the degree to which a book matches his interests/preferences), which are seldom, if ever, made by K-3 readers, and which K3Rec does not rely on.

3.3 Ranking Candidate Books

Having determined the appropriate readability level of each candidate book CB (defined by using Equation 1) and quantified the properties of CB applicable to emergent readers, K3Rec computes a single, overall *ranking score* of CB by using $CombMNZ$ [6] (as defined in Equation 7). CombMNZ, which is a popular linear combination strategy, is applied to the aforementioned scores to determine the degree to which CB (i) matches the content and illustration preferences of a reader and (ii) shows evidence of addressing book properties desirable for K-3 readers.

$$Rank(CB) = \sum_{c=1}^{7} score^c \times |score^c > 0| \qquad (7)$$

where $score^c$ is the (normalized) value of one of the scores computed in Section 3.2 and $|score^c > 0|$ 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 aggregation strategy adopted by K3Rec accounts for the fact that not all candidate books are assigned a non-zero score for each of the measures computed in Section 3.2, i.e., $Csim(CB, B)$, $Isim(CB, B)$, $Diff(CB)$, $LC(CB, B)$, $Len(CB)$, $WTS(CB)$, and $Rate(CB)$. The joint ranking considers the *strength* of each evidence regardless whether any evidence yields a zero value, as opposed to simply positioning higher in the ranking candidate books with non-zero scores for all the measures. After the joint ranking score has been computed for each candidate book, the top-3 highest-ranked books are suggested to R.

EXAMPLE 4. Consider a reader R who has read and enjoyed "Too Princessy!" by Jean Reidy, i.e., $Book_R$ as shown in Figure 3. By performing a multi-dimensional analysis on $Book_R$ using books in the BiblioNasium dataset, K3Rec suggests "Too Purpley!" by Jean Reidy ($Book_1$), "Birdie Plays Dress-Up" by Sujean Rim ($Book_2$), and "Wacky Wednesday" by Dr. Seuss ($Book_3$) in the dataset in the respective order. We have manually verified that the suggestions are relevant recommendations for R, not only because their grade levels correlate with the reading ability of R, which is at the 1.4 grade level, but also because they share similar content, have similar illustrations, and are highly-regarded and relatively-short books (in terms of their ratings and page counts, respectively) that include simple and direct narratives and address topics (i.e., LCSH such as "Stories in rhyme", "Play", and "Pictorial books") suitable for K-3 readers. □

[11]Popular book-related sites, such as Amazon.com, GoodReads.com, or Kidsread.com, also archive ratings on books.

Figure 3: Top-3 recommendations generated by K3Rec based on the interest of the reader R on the book "Too Princessy" by Jean Reidy

4. EXPERIMENTAL RESULTS

In this section, we first introduce our evaluation framework (in Section 4.1). Hereafter, we present the results of the empirical studies conducted to assess the performance of K3Rec (in Sections 4.2 and 4.3).

4.1 Evaluation Framework

Although the BookCrossing dataset[12] has been employed to evaluate book recommenders tailored to a general audience, it is not specifically designed for assessing the performance of book recommenders for emergent readers. We conducted a number of empirical studies, presented in Sections 4.2 and 4.3, on their respective dataset to validate the effectiveness of K3Rec.

The first empirical study relies on data from BiblioNasium.com, a bookmarking site set up exclusively to encourage children and teenagers to read. The BiblioNasium dataset consists of 1,705 K-3 users and their bookmarks, i.e., books assigned to the respective "bookshelves" by each of the users. The second empirical study depends on data collected using Amazon's Mechanical Turk (https://www.mturk.com/mturk/welcome), which is 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 and is being used to collect user feedback on various information retrieval designs.

Regardless of the study, the current implementation of K3Rec uses close to 20,000 books available at BiblioNasium.com as its book repository. Note, however, that besides BiblioNasium, any other book repository, such as OpenLibrary.org, can also be employed by K3Rec to make recommendations. Furthermore, as the design methodology of K3Rec relies on *topical*, *brief content*, and *writing style* descriptions, in addition to *covers*, predicted *grade levels*, *page counts*, and *ratings* of books, we retrieved the brief book descriptions, LCSH, and page count from the Library of Congress, their ratings and covers from Google Books, their writing style descriptions from book reviews (available at reputable book-related websites) using ABET, and their readability levels using TRoLL.

It is worth mentioning that the *statistical significance* of the results presented in the following sections were determined using the Wilcoxon signed-ranked test.

4.2 Evaluation of K3Rec Versus BReK12

Using the BiblioNasium dataset, we conducted an evaluation on the performance of K3Rec, which we compared with

[12]Informatik.unifreiburg.de/~cziegler/BX

the performance of BReK12 (as introduced in Section 2). We compared K3Rec with BReK12, since to the best of our knowledge BReK12 is the only existing recommender that explicitly considers the readability level of its users in making personalized book recommendations. Furthermore, we excluded other state-of-the-art approaches for (book) recommendations for comparison purpose, since (as stated in Section 2) they require *personal ratings* on books provided by individual users which are neither available for K-3 readers nor are included in the BiblioNasium dataset.

We assessed the performance of K3Rec and BReK12 using two metrics: *Mean Reciprocal Rank* (MRR) and *Normalized Discounted Cumulative Gain* (nDCG). While MRR computes the *average ranking position* of the *first* relevant book suggested by a recommender, nDCG determines the *overall* (ranking) performance of the recommender and penalizes relevant books positioned 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. To compute the aforementioned metrics, given a reader R in the BiblioNasium dataset, we treated one of his/her bookmarked books B as a book "of interest" to R. Hereafter, a book suggested to R by a recommender is treated as *relevant* to R if it is one of the remaining bookmarks of R, and is *non-relevant* otherwise, which is a commonly-employed evaluation protocol. (This evaluation is repeated for each of R's bookmarks.) Since only books that have been bookmarked by a user are considered *relevant*, it is not possible to account for potentially relevant books a user has not bookmarked, which is a well-known limitation of this evaluation protocol. As the limitation applies to both BReK12 and K3Rec, the results of the empirical studies are consistent for the comparison purpose.

As shown in Figure 4, K3Rec achieves a significant improvement ($p < 0.001$) over BReK12 in terms of nDCG, which are 0.79 and 0.65, respectively. Moreover, according to the computed MRR scores, users of K3Rec are expected to browse, on the average, *one* ($= \frac{1}{0.77} = 1.2$) book suggestion before locating a relevant one, as opposed to BReK12 users who are required to browse through *two* ($= \frac{1}{0.62} = 1.6$) before a relevant book is located. The difference in MRR between the recommenders is statistically significant ($p < 0.001$). The experimental results verify not only the effectiveness of K3Rec in applying its the multi-dimensional recommendation strategy, but also the choice of using book meta-data, instead of bookmarks on books used by BReK12. Unlike bookmark data created by more mature readers, bookmarks are rarely created by K-3 readers.

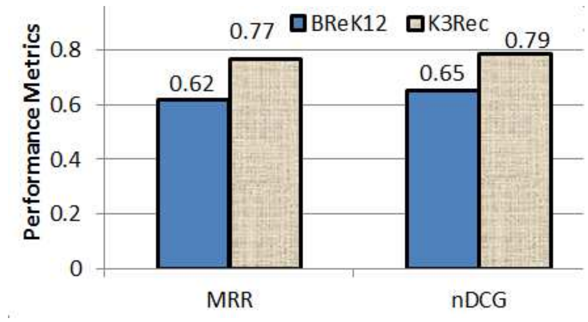

Figure 4: Performance evaluation of BReK12 and K3Rec using the BiblioNasium dataset

4.3 Mechanical Turk Appraisers

To further assess the performance of K3Rec, we conducted a survey using Mechanical Turk appraisers[13] who identified, among a provided set of three books (generated using K3Rec), the ones that relate to a given book B. The purpose of this survey is to emulate the behavior of K3Rec when presented with B, and quantify the degree of relevance of the generated suggestions based on the opinion of independent appraisers. This survey quantifies the degree of relevant suggestions made by K3Rec based on the opinions of independent appraisers.

We created ten HITs (Human Intelligent Task) on Mechanical Turk, each with a different book and its corresponding set of suggestions made by K3Rec. (A sample HIT is shown in Figure 5.) We collected responses to the HITs from 400 independent appraisers during the month of April 2014. The responses provided by each appraiser are treated as the "gold standard", i.e., the chosen books are treated as *relevant* to the given book in the corresponding HIT.

The accuracy ratios computed using the collected responses, which reflect the proportion of books treated as relevant by independent appraisers among the top-3 books included in each HIT, are shown in Figure 6. Among the appraisers who provided their occupation, 63% were teachers, parents of young readers, or librarians. Given that (i) parents/teachers/ librarians are the ones who often select books for K-3 readers and (ii) the impossibility of directly interacting with K-3 readers using Mechanical Turk, it is appropriate to quantify the performance of K3Rec reflected by the opinions of librarians, parents of young readers, and teachers separately from other appraisers with diverse occupations/professions. As shown in Figure 6, the accuracy ratios calculated according to parents/teachers/librarians responses yield a statistically significant improvement ($p < 0.001$) over the one based on all the collected responses. The results compiled using the opinion of "experts," i.e., parents/teachers/librarians, in books targeting emergent readers are of special importance in assessing the performance of K3Rec, given the lack of benchmark datasets to evaluate recommendation tools for K-3 readers. Moreover, the fact that appraisers who are "experts" appreciate the recommendations made by K3Rec more than general appraisers provides further evidence of the usefulness of K3Rec in suggesting books for K-3 readers in locating suitable reading materials. Based on the feedback collected through Mechanical Turk, we have observed that consistently, almost 2 out of the 3 generated book recommendations were treated as relevant by Mechanical turk appraisers, which demonstrates the effectiveness of K3Rec in locating books suitable for emergent readers.

To evaluate the degree to which books recommended by K3Rec are preferred over those suggested by recommendation modules at well-known book-related websites, we created another set of 10 HITS using Mechanical Turk. We have selected several well-known recommenders that adopt diverse strategies in making book suggestions: (i) Ama-

[13]We are aware that crowdsourcing assessments can be affected by spam. To address this issue, we included in each of our surveys a book that did not align with the task in the survey. Appraisers that selected said book were treated as spammers and their assessment discarded.

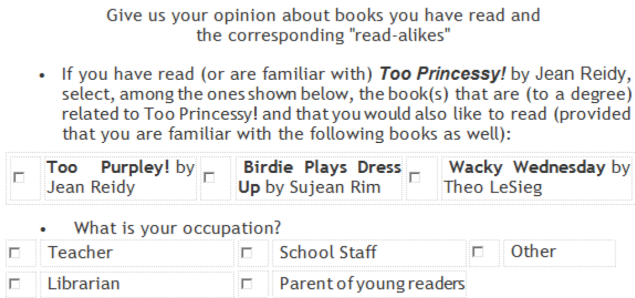

Figure 5: A sample survey conducted on Mechanical Turk to determine the relevance of K3Rec-generated recommendations

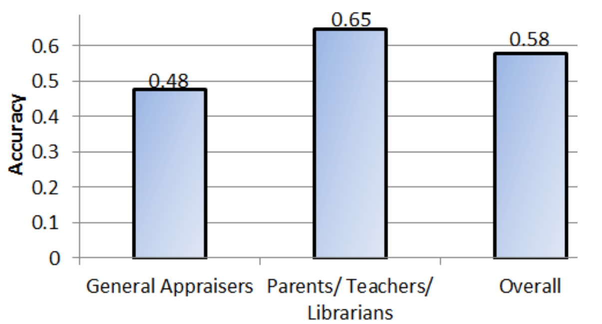

Figure 6: Performance evaluation of K3Rec-generated recommendations based on the opinions of parents/teachers/librarians and other appraisers

zon, which considers purchasing patterns of its users [14], GoodReads, [14] which "combines multiple proprietary algorithms that analyze 20 billion data points to better predict which books people want to read next", and (iii) NoveList,[15] which examines a number of book-related information, such as title and publication date, for recommending books.

Each HIT (see Figure 7 for a sample) included the top-2 recommendations (in which some of them are identical) made by NoveList, GoodReads, Amazon, and K3Rec for a given sample book B, respectively. Appraisers were asked to select the top-two books most closely related to B, which were treated as the *gold standard* for B.

Based on the 400 responses collected during the month of April 2014, we computed the accuracy of the top-2 recommendations made by K3Rec and each of the recommenders considered for comparison purpose. As shown in Figure 8, recommendations made by K3Rec and Amazon are preferred over the suggestions made by GoodReads and NoveList. Furthermore, the improvement, in terms of accuracy ratios, achieved by K3Rec over GoodReads and NoveList is statistically significant ($p < 0.001$).

In terms of the overall accuracy, K3Rec outperforms Amazon ($p < 0.05$). While K3Rec considers books provided directly by K-3 readers (or their parents/teachers) to generate personalized suggestions, recommendations made by Amazon that target children are the results of extensive analysis of the purchasing patterns of adults, which might not accurately reflect the direct interests/preferences of emergent readers in books. More importantly, K3Rec can treat a book K as a candidate suggestion immediately after K is published, unlike Amazon which requires a number of purchasing transactions involving K to recommend it.

[14] http://goo.gl/AZ8xvv

[15] support.epnet.com/knowledge_base/detail.php?id=4772

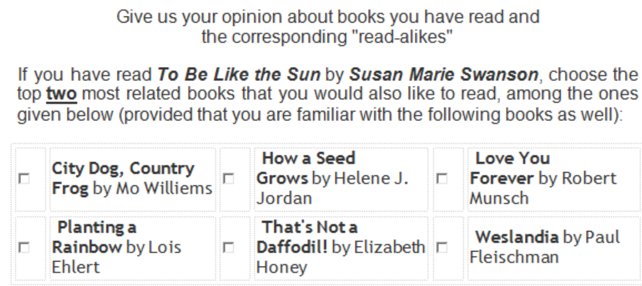

Figure 7: A Mechanical Turk HIT on the book "To Be Like the Sun"

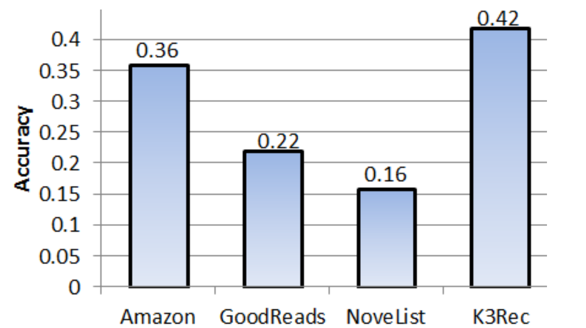

Figure 8: Accuracy achieved by Amazon, GoodReads, NoveList, and K3Rec based on the opinions of Mechanical Turk appraisers

5. CONCLUSIONS AND FUTURE WORK

We have presented K3Rec, an unsupervised book recommender developed for K-3 readers who are not currently targeted by existing recommenders. K-3 readers are an essential audience, given that individuals' reading habits are developed early in life. Unlike current state-of-the art recommenders, K3Rec does not rely on personal social media data, such as personal ratings or bookmarks, which are rarely created by emergent readers, to make recommendations. Instead, K3Rec takes advantage of publicly-available (meta)data on books and (i) examines properties of books that target young audiences, such as their short length and simple and direct writing style, (ii) considers the suitability of topics addressed in books, (iii) analyzes books' contents, and (iv) compares book illustrations, which offer children joy in reading while at the same time help them develop visual thinking skills. The design goal of K3Rec is to assist K-3 readers, their parents, and teachers in their quest for books, either for pleasure reading or knowledge acquisition. K3Rec enriches its readers' choices on books and encourages them to read so that they could become lifelong readers. We have conducted empirical studies using data from BiblioNasium to validate the effectiveness of K3Rec and its superiority over existing recommenders that explicitly consider the reading ability of its users. Conducted experiments using a crowdsourcing platform have further verified the relevance of books suggested by K3Rec, which outperforms the recommenders at Amazon, GoodReads, and NoveList.

For future work, we would like to extend the performance evaluation on K3Rec to determine the impact K3Rec has on the reading and learning habits of emergent readers. Furthermore, we would like to enhance the functionality of K3Rec by examining existing image-matching models and, if necessary, develop one that would allow us to perform a more in-depth examination of book illustrations to distinguish, for example, *a little girl from a doll*. In doing so, we anticipate that more relevant book suggestions could be generated, which will improve the effectiveness of our proposed recommender.

6. REFERENCES

[1] R. Allington and E. Gabriel. Every Child, Every Day. *Reading: The Core Skill*, 69(6):10–15, 2012.

[2] R. Benjamin. Reconstructing Readability: Recent Developments and Recommendations in the Analysis of Text Difficulty. *Educational Psychology*, 24:63–88, 2012.

[3] A. Budanitsky and G. Hirst. Evaluating WordNet-based Measures of Lexical Semantic Relatedness. *Computational Linguistics*, 32(1):13–47, 2006.

[4] The Annie E. Casey Foundation. Early Warning Confirmed: A Research Update on Third-Grade Reading. Available at http://goo.gl/HQrPOA, 2013.

[5] C. Coulter and M. Smith. The Construction Zone: Literary Elements in Narrative Research. *Educational Researcher*, 38(8):577–590, 2009.

[6] W. Croft, D. Metzler, and T. Strohman. *Search Engines: Information Retrieval in Practice*. Addison Wesley, 2010.

[7] R. Dong, M. O'Mahony, M. Schaal, K. McCarthy, and B. Smith. Sentimental Product Recommendation. In *Proceedings of ACM conference on Recommender systems (RecSys)*, pages 411–414, 2013.

[8] I. Fountas and G. Pinnell. *Matching Books to Readers: Using Leveled Books in Guided Reading, K-3*. Heinemann, 1999.

[9] L. Girard, L. Girolametto, E. Weitzman, and J. Greenber. Educators' Literacy Practices in Two Emergent Literacy Contexts. *Journal of Research in Childhood Education*, 27(1):46–60, 2013.

[10] S. Givon and V. Lavrenko. Predicting Social-Tags for Cold Start Book Recommendations. In *Proceedings of ACM conference on Recommender systems (RecSys)*, pages 333–336, 2009.

[11] C. Juel. Learning to Read and Write: A Longitudinal Study of Fifty-Four Children from First Through Fourth Grade. *Journal of Educational Psychology*, 80:437–447, 1988.

[12] L. Justic and J. Kaderavek. Using Shared Storybook Reading to Promote Emergent Literacy. *Teaching Exceptional Children*, 34(4):8–13, 2002.

[13] J. Koberstein and Y.-K. Ng. Using Word Clusters to Detect Similar Web Documents. In *Proceedings of Second International Conference on Knowledge Science, Engineering, and Management (KSEM 2007)*, pages 215–228, 2006.

[14] G. Linden, B. Smith, and J. York. Amazon.com Recommendations: Item-to-item Collaborative Filtering. *IEEE Internet Computing*, 7(1):76–80, 2003.

[15] H. Mai and M. Kim. Utilizing Similarity Relationships Among Existing Data for High Accuracy Processing of Content-Based Image Retrieval. *Multimedia Tools and Applications*, January:1–30, 2013.

[16] N. Manouselis, H. Drachsler, K. Verbert, and E. Duval. *Recommender Systems for Learning*. Springer Briefs in Electrical and Computer Engineering, 2013.

[17] J. Oakhill and K. Cain. The Precursors of Reading Ability in Young Readers: Evidence from a Four-Year Longitudinal Study. *Scientific Studies of Reading*, 16(2):91–121, 2012.

[18] Ministry of Education of Ontario. A Guide to Effective Instruction in Reading, Kindergarten to Grade 3. Available at http://goo.gl/UCo5e3, 2005.

[19] M. Pera and Y.-K. Ng. What to Read Next?: Making Personalized Book Recommendations for K-12 Users. In *Proceedings of ACM conference on Recommender systems (RecSys)*, pages 113–120, 2013.

[20] M. Pera. *Using Online Data Sources to Make Recommendations on Reading Materials for K-12 and Advanced Readers*. PhD Dissertation, Brigham Young University, April 2014.

[21] M. Renck. *Young Children and Picture Books (2nd Ed.)*. National Association for the Education of Young Children, 2004.

[22] C. Robinson, J. Larsen, J. Haupt, and J. Mohlman. Picture Book Selection Behaviors of Emergent Readers: Influence of Genre, Familiarity, and Book Attributes. *Reading Research and Instruction*, 36(4):287–304, 1997.

[23] K. Roskos, J. Christie, and D. Richgels. The Essentials of Early Literacy Instruction. *Young Children*, 58(2):52–60, 2003.

[24] J. Saricks. *Readers' Advisory Service in the Public Library, 3^{rd} Ed.* ALA Store, 2005.

[25] A. Sieg, B. Mobasher, and R. Burke. Improving the Effectiveness of Collaborative Recommendation with Ontology-based User Profiles. In *Proceedings of International Workshop on Information Heterogeneity and Fusion in Recommender Systems (HetRec 2010)*, pages 39–46, 2010.

[26] S. Vanneman. Keep Them Reading. *School Library Monthly*, 27(3):21–22, 2010.

[27] G. Whitehurst and C. Lonigan. *Handbook of Early Literacy Research, Volume 1*, chapter Emergent Literacy: Development from Prereaders to Readers. The Guilford Press, 2003.

[28] C. Yang, B. Wei, J. Wu, Y. Zhang, and L. Zhang. CARES: A Ranking-oriented CADAL Recommender System. In *Proceedings of ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, pages 203–212, 2009.

[29] K. Yi and L. Chan. Revisiting the Syntactical and Structural Analysis of Library of Congress Subject Headings for the Digital Environment. *Journal of the Association for Information Science and Technology (JASIST)*, 61(4):677–687, 2010.